DAHLGREN DIVISION
NAVAL SURFACE WARFARE CENTER
Dahlgren, Virginia 22448-5100

NSWCDD/TR-95/123

MULTISET APPROACH TO DISCRIMINATION OF SPATIAL STATISTICS

GEORGE W. ROGERS
SYSTEMS RESEARCH AND TECHNOLOGY DEPARTMENT

CAREY E. PRIEBE
DEPARTMENT OF MATHEMATICAL SCIENCES
JOHNS HOPKINS UNIVERSITY

OCTOBER 1995

REPORT DOCUMENTATION PAGE

Form Approved
OBM No. 0704-0188

1. AGENCY USE ONLY (*Leave blank*)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	October 1995	

4. TITLE AND SUBTITLE

Multiset Approach to Discrimination of Spatial Statistics

5. FUNDING NUMBERS

6. AUTHOR(s)

George W. Rogers, Carey E. Priebe

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Surface Warfare Center
Dahlgren Division (Code B10)
17320 Dahlgren Rd.
Dahlgren, VA 22448-5100

8. PERFORMING ORGANIZATION REPORT NUMBER

NSWCDD/TR-95/123

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)

The need to discriminate and classify spatial patterns is central to pattern recognition. The emergence of the field of computational statistics, made possible by recent advances in digital computing, has opened up a new way of approaching pattern recognition. In particular, it is no longer necessary to assume restrictive statistical models for spatially correlated data. It thus becomes important to develop model-free approaches to the description of spatial patterns.

14. SUBJECT TERMS

pattern recognition, spatial patterns, multiset

15. NUMBER OF PAGES

16

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

FOREWORD

The need to discriminate and classify spatial patterns is central to pattern recognition. The emergence of the field of computational statistics, made possible by recent advances in digital computing, has opened up a new way of approaching pattern recognition. In particular, it is no longer necessary to assume restrictive statistical models for spatially correlated data. It thus becomes important to develop model-free approaches to the description of spatial patterns.

This work was supported in part by the Office of Naval Research (R&T No. 4424314) and the Naval Surface Warfare Center, In-house Laboratory Independent Research Program.

This report was reviewed by Dr. Richard Lorey, Head, Advanced Computation Technology Group.

Approved by:

MARY E. LACEY, Head
Systems Research and Technology Department

INTRODUCTION

Consider the problem of detecting and classifying homogeneous regions in a spatially correlated signal such as an image. One approach has been to assume some model describes the spatial process and to fit the model parameters. As an example, consider the model for a one-dimensional simultaneous autoregressive (SAR) process[1,2]

$$z_i = \mu_i + \sum_j K_{ij}(z_j - \mu_j) + \varepsilon_i \tag{1}$$

where z_i is signal value at bin i with mean value μ_i and with independent, identically distributed (iid) noise ε_i (usually assumed normal). The weights K_{ij} specify the neighborhood and weighting of the spatial dependence. To use this model, a particular form of the weights must be chosen and a parametric form for the iid noise is assumed. Fitting the model to the signal then consists of determining the parameters of the noise model. For Gaussian noise, the variance of the noise is the unknown parameter.

This approach has the drawbacks of the need to assume a parametric model, the limited number of models for which parameter estimators can be derived, and the limited ability of the resultant model to accurately describe a wide range of correlated phenomena. While other models of spatial correlation abound,[1,2] they suffer from analagous drawbacks.

An alternative approach is to treat the correlated observations as if they were independent or uncorrelated and proceed to perform nonparametric or semiparametric density estimates under this assumption. A good example of this approach as applied to fractal dimension analysis is given in Reference 3, while an application to change point detection can be found in Reference 4. This latter approach has demonstrated significant promise for the discrimination of different spatial processes even when no simple parametric model such as Equation (1) adequately describes the processes.

When applying this alternative approach, it is desirable to know explicitly what assumptions are being made, and what the ramifications of those assumptions may be. It is the purpose of this report to examine and make explicit the required assumptions as well as discuss some of their ramifications.

In the next section, a sequence of mappings is developed for discrimination of spatial processes. This is followed by a section devoted to the relationship between these mappings and the set of probability density functions (pdf's). The final section presents some concluding remarks.

MULTISET MAPPINGS AND ERGODICITY

Nonparametric density estimation techniques abound for iid data. Examples[5] of nonparametric density estimators include histograms, frequency polygons, average shifted histograms, and kernel estimators. Semiparametric density estimators are based on the notion of retaining the nonparametric flexibility of the nonparametric techniques while limiting the growth rate in the model complexity. The adaptive mixture model[6] is a good example. It is a parametric mixture model that can add terms (and hence complexity) in a data-driven manner. It thus does not have a fixed number of parameters and hence cannot be properly termed parametric. Since its complexity grows at a much slower rate than that of the nonparametric techniques, a suitable term is semiparametric.

Thus, on the one hand there is a powerful set of density estimation tools that do not make any model assumptions but require iid data, while on the other hand the problem under consideration involves spatially correlated data that is manifestly non-iid. This presents a conundrum on how to proceed. At first glance, the choices are to either scrap the nonparametric tools when the data is correlated, or to make the false assumption that the correlations do not exist.

An alternative that permits the use of the nonparametric tools while not making the iid assumption blindly is to develop a sequence of mappings from the correlated data to data sets where succeeding sets in the sequence (may) retain more and more of the spatial information.

Consider only spatial processes on a lattice. Let D represent the index set for the lattice points. Thus for a two-dimensional $n \times n$ lattice, $D=\{(i,j) \mid i=1,...,n;\ j=1,...,n\}$. Then the random process can be written as

$$\{Z(s) \mid s \in D\}\ , \tag{2}$$

with a particular realization of this process being denoted by

$$\{z(s) \mid s \in D\}\ . \tag{3}$$

The random process Equation (2) is usually defined[1] through the finite-dimensional distributions

$$F_{s_1,\ldots,s_m}(z_1,\ldots,z_m) = P\{Z(s_1) \le z_1,\ldots,Z(s_m) \le z_m\}, \qquad m \ge 1\ . \tag{4}$$

This in turn leads to the consideration of the marginal pdf's (provided they exist)

$$f(z_i) = \frac{d}{dz_i} P(Z(s_i) \le z_i) \tag{5}$$

as well as joint pdf's, a two-dimensional example of which is

$$f(z_i, z_j) = \frac{d\ d}{dz_j dz_i} P(Z(s_i) \le z_i, Z(s_j) \le z_j) . \tag{6}$$

If there are a large number of realizations of the same process, then it is an option to estimate the density function separately at each lattice site. If however, there is only one realization and the interest is in characterizing the process in terms of densities, assumptions must be made to proceed. One logical assumption is that of strong (or strict) stationarity of the process, defined by the conditions that

$$F_{s_1+h, \ldots, s_m+h}(z_1, \ldots, z_m) \equiv F_{s_1, \ldots, s_m}(z_1, \ldots, z_m) \tag{7}$$

for all $(m \ge 1)$ and all $(s_j + h) \in D$. Under this assumption,

$$f(z_i) \equiv f(z_j) \qquad \forall (s_i, s_j \in D) , \tag{8}$$

with similar identities holding for all possible joint pdf's.

Next suppose that it is desired to characterize a strongly stationary process by its marginal pdf based on a single realization of the process. The estimate can be based on the set of values from this single realization. In this case, the empirical cumulative distribution function (ecdf), based on n elements in D is

$$F_n(z) = \frac{\#\{z_i \le z\}}{n} . \tag{9}$$

For the ecdf to converge to the true cumulative distribution function (cdf), the observations must be identically distributed (guaranteed by the assumption of strong stationarity) and the number of observations must go to infinity. Additionally, the usual requirement is that the observations are independent as well as identically distributed. Since the observations are in general not independent, an alternative requirement is that

$$\lim_{n \to \infty} F_n(z) = F(z) \tag{10}$$

where

$$n \to \infty \Leftrightarrow \#s_i \in D \to \infty, \tag{11}$$

and the limit is the same for almost every realization of the process. This requirement is just the requirement of ergodicity.[1]

In practice, data sets are always finite corresponding to a finite number of elements in the index set D. Define the following mapping that simply maps a particular realization of a process Equation (2) to a set of values where in general each element in the new set is indexed by the number of times it occurred in the realization. Formally,

$$M_0(z(s) \,|\, s \in D) \to y; \tag{12}$$

that is, each (site indexed) value of the realization of the process is mapped to simply its value, while the map operating on the set gives

$$M_0(\{z(s) \,|\, s \in D\}) \to \{y^{(n_y)} \,|\, n_y = \#(z(s) = y)\}. \tag{13}$$

This last set is formally a multiset as there is an associated value (the number index) with each element. The formal requirement of a multiset representation is only needed when the same value can occur in the process with nonzero probability.

Under the assumption of ergodicity, a pdf and/or cdf may be estimated by any of the methods normally employed with iid data.

The map definition can be extended to yield multivariate sets as follows. Let

$$\begin{aligned}
&M_h(z(s), z(s+h) \,|\, s, s+h \in D) \to (y_1, y_2) \\
&M_{h_1, h_2}(z(s), z(s+h_1), z(s+h_2) \,|\, s, s+h_1, s+h_2 \in D) \to (y_1, y_2, y_3) \\
&\quad \bullet \\
&\quad \bullet \\
&\quad \bullet \\
&M_{h_1, \ldots, h_m}(z(s), \ldots, z(s+h_m) \,|\, s, \ldots, s+h_m \in D) \to (y_1, \ldots, y_m)
\end{aligned} \tag{14}$$

so that, operating on the set of process values,

$$M_h(\{z(s)\,|s\in D\}) \rightarrow \{(y_1,y_2)^{(n_y)}\,|n_y = \#((z(s)=y_1)\wedge(z(s+h)=y_2))\}$$

$$M_{h_1,h_2}(\{z(s)\,|s\in D\}) \rightarrow \left\{(y_1,y_2,y_3)^{(n_y)}\,|n_y = \#\left(\begin{array}{c}(z(s)=y_1)\\ \wedge\,(z(s+h_1)=y_2)\\ \wedge\,(z(s+h_2)=y_3)\end{array}\right)\right\}$$

•
•
•

$$M_{h_1,\ldots,h_{m-1}}(\{z(s)\,|s\in D\}) \rightarrow \left\{(y_1,\ldots,y_m)^{(n_y)}\,|n_y = \#\left(\begin{array}{c}(z(s)=y_1)\\ \wedge\ldots\\ \wedge\,(z(s+h_{m-1})=y_m)\end{array}\right)\right\} \tag{15}$$

This sequence of mappings can be augmented by considering linear combinations of the random variables Z(s) corresponding to performing convolutions or weighted sums. Let this derived process be denoted by

$$\{X(s)\,|s\in D\}\,, \tag{16}$$

with a particular realization of this process being denoted by

$$\{x(s)\,|s\in D\}\,. \tag{17}$$

Then all of the mappings can be used with $x \rightarrow z$. As an additional alternative, consider mixed mappings where for example,

$$M_0(z(s),x(s)\,|s\in D) \rightarrow (y_1,y_2)$$

$$M_0(\{(z(s),x(s))\,|s\in D\}) \rightarrow \left\{(y_1,y_2)^{(n_y)}\,|n_y = \#\left(\begin{array}{c}(z(s)=y_1)\\ \wedge\,(x(s)=y_2)\end{array}\right)\right\}. \tag{18}$$

This type of mapping is suggested by the SAR process Equation (1) where a joint density might be of interest

$$f(z(s_i),x(s_i)) = f\left(z(s_i),\sum_k z(s_k)\right). \tag{19}$$

PDFs CORRESPONDING TO MULTISET MAPPINGS

In the limit as the lattice size goes to infinity, each multi-set mapping corresponds to a density if the spatial process is strongly stationary. (It may or may not correspond to a density under weaker conditions.) Under any given multiset mapping, a many-to-one mapping can exist between the set of all ergodic spatial processes and the set of pdf's (including the null element). This serves to produce a set of equivalence classes of ergodic spatial processes under each mapping. This can be seen pictorially in Figure 1, which shows set relationships under a multiset mapping M_k in the limit of an infinite number of observations. Elements of the set of spatial statistics undergo a many-to-one mapping to a set of equivalence classes. Each equivalence class corresponds to a unique iid (potentially multivariate or joint) pdf. One of the equivalence classes corresponds to the non-existance of a pdf.

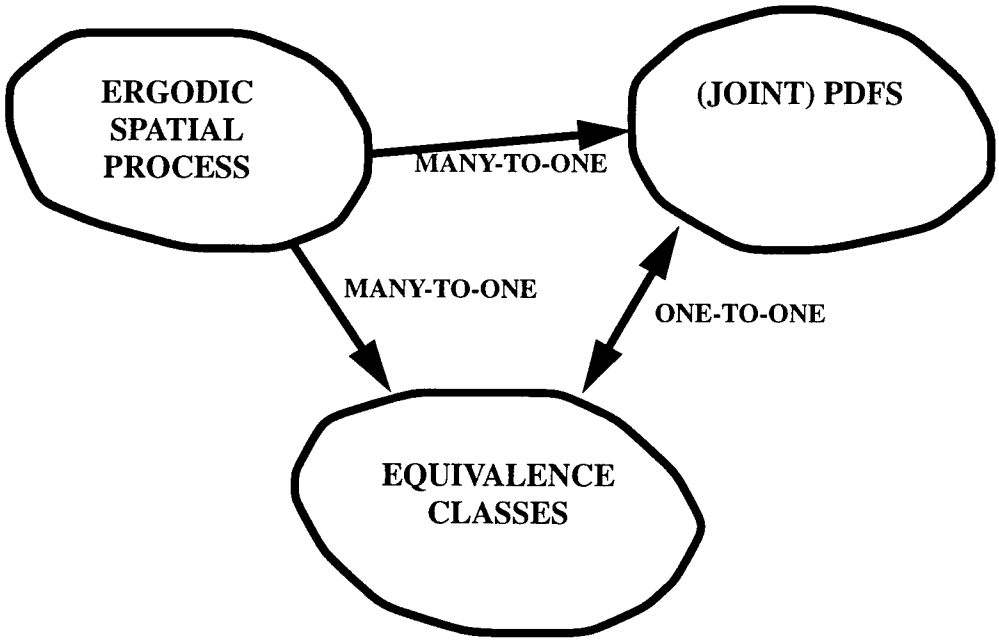

FIGURE 1. SET RELATIONSHIPS UNDER A MULTISET MAPPING M_k

Next, consider what happens when these assumptions are applied to a process that is not ergodic or even stationary. Asymptotically, the pdf corresponding to a mapping may or may not exist. Alternatively, there is the case where different realizations of a spatial process (even in the limit of an infinite lattice) converge to different densities. The former case, where no pdf exists, can be handled by augmenting the set of pdf's with an element corresponding to a null pdf. The latter case amounts to the process having a many-to-many correspondence rather than the many-to-one correspondence shown in Figure 1.

Consider the following example depicted in Figure 2. Figure 2a shows two checkerboard patterns with different scales. In Figure 2b, the two patterns (in the limit of infinite spatial extent) map to the same pdf under M_0. Figure 2c uses a discrete one-dimensional representation of the two-dimensional joint densities resulting from the multiset map $M_{(1,0)}$. The two elements of the ordered pairs correspond to the independent coordinates of the two-dimensional joint pdf. It has been mapped to a single axis here for simplicity. Additionally, any binary pattern with equal numbers of pixels with values of one and zero will fall into the same equivalence class as the checkerboard patterns under M_0. Any patterns with unequal numbers of on and off pixels or with intermediate values will separate into different equivalence classes under the M_0 map.

Thus it is seen that even though the M_0 map discards all ordering information, a significant amount of information remains that can be used via pdf's to group spatial processes into equivalence classes. If it is desired to discriminate between two spatial processes that map to the same equivalence class under M_0, it suffices to find another element in the multi-set map sequence for which discrimination will occur.

The preceeding example was chosen as a simple example to illustrate the concepts that have been presented. The more usual application involves a continuous valued function at each lattice point.

In this case, the pdf is continuous and finite sample size becomes an issue. While all of the nonparametric density estimators mentioned earlier are designed for use with continuous densities, all possess bias and there will exist variance due to finite sample size. In this case an exact match between densities cannot be required because of the error in estimating them. Rather, the classification of the finite samples must be based on a measure of the difference between density estimates. Examples of such a measure include L_1, L_2, and Kullback-Leibler distance. Because of bias and variance in the density estimates, it is necessary to set a threshold for grouping density estimates (and hence data sets) into equivalence classes. Unfortunately, an appropriate threshold is problem and sample size dependent and will not be addressed in this report.

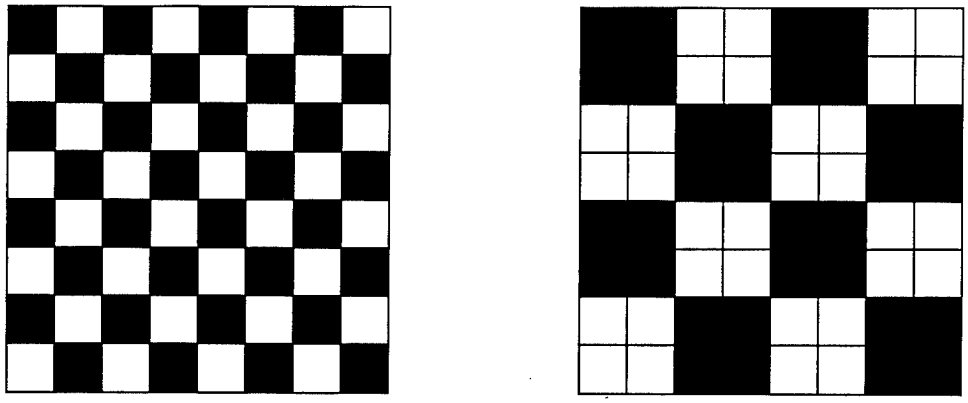

FIGURE 2A. CHECKERBOARD PATTERNS AT TWO DIFFERENT SCALES

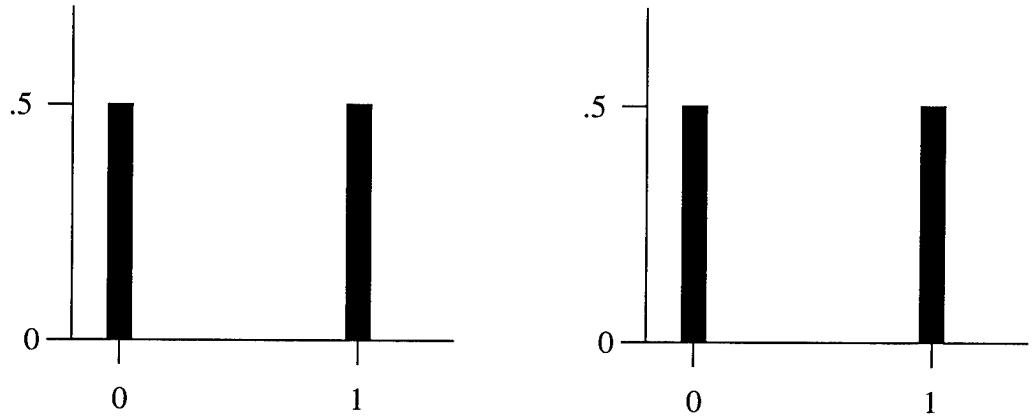

FIGURE 2B. DISCRETE PROBABILITY DENSITY FUNCTIONS CORRESPONDING TO THE TWO CHECKERBOARD PATTERNS UNDER THE MULTI-SET MAP M_0

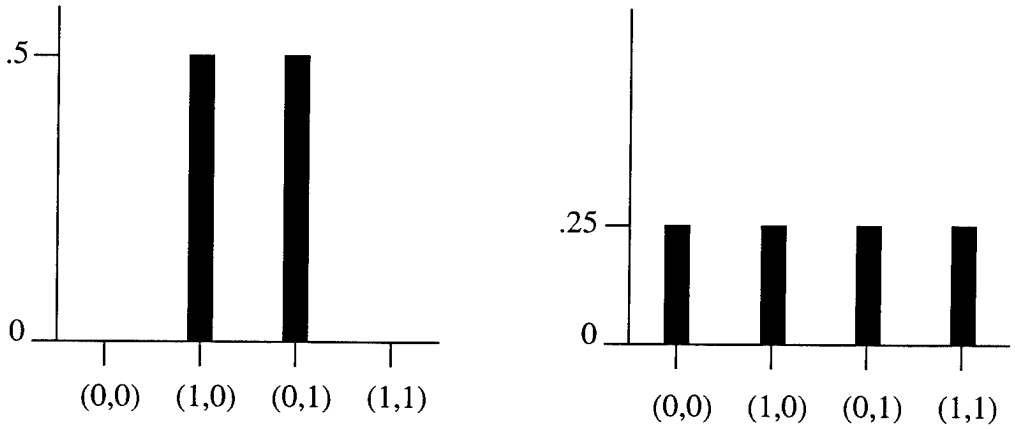

FIGURE 2C. DISCRETE ONE-DIMENSIONAL REPRESENTATION OF THE TWO-DIMENSIONAL JOINT DENSITIES RESULTING FROM THE MULTISET MAP $M_{(1,0)}$

CONCLUSIONS

An approach to characterizing spatial statistics has been presented that is based on using one or more mappings that convert spatially correlated values to a form, which allows the use of nonparametric and semiparametric density estimation techniques to characterize the spatial statistics through the density estimate of the mapped values. Since nonparametric and semiparametric density estimation techniques typically are based on the iid assumption, they must be used on data sets that are iid or else on data that behaves as if it is iid asymptotically. The mappings that have been proposed in effect discard any spatial correlation information and result in multisets composed of n-tuples of distinct values with associated indices that count the number of occurances of that distinct n-tuple set of values. These multisets have no explicit or implicit ordering of the n-tuples and hence possess an asymptotic equivalence with n-dimensional joint pdf's if the spatial process that generated them is ergodic.

Using the multiset mappings corresponds to asking the question of which asymptotically equivalent iid density does the spatial process in question correspond to under a particular multiset map. For ergodic spatial processes this correspondence is well defined. As was seen in the example, a many-to-one mapping exists for ergodic spatial processes to an asymptotically equivalent density. This serves to divide spatial processes into equivalence classes that will differ under different mappings.

Thus, this approach can be viewed as one with the intent to discriminate between particular types of spatial processes (using nonparametric/semiparametric tools) rather than to directly describe the spatial processes in terms of specific models.

Finally, to place this approach in perspective, consider these quotes from Reference 1.

> "Ergodicity is an assumption made to allow inference to proceed for a series of nonindependent observations. It might only be verifiable in the sense that one fails to reject it. This should not be too worrisome because scientific discovery generally proceeds in this way."

and

> "It seems that statisticians are using only the part of the ergodicity assumption that guarantees the sample mean and covariances converge to their population counterparts."

Thus, this report started with the usual method of dealing with nonindependent observations, but rather than simply computing sample means and covariances, a broad semiparametric/nonparametric approach to characterizing a broad class of spatial processes has been formulated. This approach is based on a sequence of mappings and their resulting semiparametric/nonparametric probability density estimates, which serve to group spatial processes into sets of equivalence classes which are a function of the mapping used. While the approach could have been formulated

without explicit reference to the multiset mappings, the goal was to explicitly lay out as many of the assumptions as possible that have been implicit in earlier work.

In a companion report, an example of this approach involving continuous densities will be provided. In future work, some related issues such as finite sample size, the role of imposed measure, and the finite sample impact of nonergodicity will be examined in more detail.

REFERENCES

1. Cressie, N. A. C., *Statistics For Spatial Data*, Wiley, New York, 1993.

2. Ripley, B. D., *Spatial Statistics*, Wiley, New York, 1981.

3. Priebe, C. E.; Lorey, R. A.; Marchette, D. J.; Solka, J. L.; and Rogers, G. W., "Nonparametric spatio-temporal change point analysis for early detection in mammography," *Digital Mammography*, Ed. Gale, A. G.; Astley, S. M.; Dance, D.R.; and Cairns, A.Y., Elsevier, New York, 1994, pp. 111-120.

4. Solka, J. L.; Priebe, C. E.; and Rogers, G. W., "An initial assessment of discriminant surface complexity for power law features," *Simulation*, Vol. 58, no. 5, 1992, pp. 311-318.

5. Scott, D. W., *Multivariate Density Estimation*, Wiley, New York, 1992.

6. Priebe, C. E., "Adaptive Mixtures," *J. Amer. Stat. Soc.*, Vol. 89, pp. 796-806, 1994.

DISTRIBUTION

	COPIES
DOD ACTIVITIES (CONUS)	
ATTN HAWKINS CODE 342PS OFFICE OF NAVAL RESEARCH 800 N QUINCY ST ARLINGTON VA 22217	1
ATTN CODE E29L (TECHNICAL LIBRARY) COMMANDING OFFICER CSSDD NSWC 6703 W HIGHWAY 98 PANAMA CITY FL 32407-7001	1
DEFENSE TECHNICAL INFORMATION CTR 8725 JOHN J KINGMAN SUITE 0944 FT BELVOIR VA 22060-6218	2
NON-DOD ACTIVITIES (CONUS)	
THE CNA CORPORATION PO BOX 16268 ALEXANDRIA VA 22302-0268	1
ATTN GIFT AND EXCHANGE DIVISION LIBRARY OF CONGRESS WASHINGTON DC 20540	4
ATTN PRIEBE DEPT OF MATHEMATICAL SCIENCES JOHNS HOPKINS UNIVERSITY BALTIMORE MD 21218	10
INTERNAL	
B	1
B10 (ROGERS)	40
B10 (LOREY)	1
B10 (MARCHETTE	1
B10 (SOLKA)	1
C	1
D	1
B05 (MOORE)	1
E231	3
E282 (SWANSBURG)	1
G33 (POSTON)	1